



# **HOMOLOGOUS PROTEINS:**

## Top BLAST Hits:

	Score	E
CFA 145000030337444 /altid=gi 12017941 /def=gb AAG45335.1 AF295...	975	0.0
CFA 114000033649823 /altid=gi 10945621 /def=gb AAG24618.1 AF293...	537	e-169
CFA 150000003782430 /altid=gi 8677401 /def=gb AAF75589.2 AF1736...	531	e-168
CFA 140000002720069 /altid=gi 8248427 /def=gb AAF74195.1 AF2496...	537	e-166
CFA 37000000006802 /altid=gi 7243145 /def=dbj BAA92620.1  (AB03...	573	e-164
CFA 13000005069115 /altid=gi 5870333 /def=ref NP_006832.1  tran...	500	e-140
CFA 33000001154721 /altid=gi 7406950 /def=gb AAF61849.1 AF15985...	496	e-139
CFA 65000019404613 /altid=gi 9506337 /def=ref NP_061849.1  amin...	495	e-139
CFA 100000004435450 /altid=gi 8926332 /def=gb AAF81797.1 AF2730...	492	e-138
CFA 335001093689635 /altid=gi 11434147 /def=ref XP_006635.1  hy...	480	e-134

## EST:

gi 10934204 /dataset=dbest /taxon=96...	1072	0.0
gi 10286121 /dataset=dbest /taxon=96...	718	0.0
gi 9872634 /dataset=dbest /taxon=960...	680	0.0
gi 2606674 /dataset=dbest /taxon=9606 ...	549	e-154
gi 9882497 /dataset=dbest /taxon=960...	541	e-151
gi 689641 /dataset=dbest /taxon=9606 /...	525	e-147

## **EXPRESSION INFORMATION FOR MODULATORY USE:**

### library source:

#### Expression information from BLAST dbEST hits:

gi|10934204 Whole embryo (mainly head)  
gi|10286121 Hepatocellular carcinoma  
gi|9872634 Non-cancerous liver  
gi|2606674 fetal liver spleen  
gi|9882497 Non cancerous liver  
gi|689641 Liver

#### Expression information from PCR-based tissue screening panels:

Mixed tissue (Brain, Heart, Kidney, Lung, Spleen, Testis, Leukocyte)

1 MDPMELHIVN IEPDDESSSG ESAPDSYIRI GNSEKAAMSS QFANEDTESQ  
 51 KFLTNFPLGK KKLADYADEH HPGTTSFGMS SFNLSNAIMG SSILGLSYAM  
 101 AYTGVILFIH MLLAVAILSL YSVHLLKTA KEGGSLIYEK LGEKAFGWPG  
 151 KIGAFVSTIM QNIGAMSSYL FIIHVELPEV IRAFMGLEEN TGEWYLNQNY  
 201 LKIFVAVGII LPLSLLKNLG YLGYTSGFSL TCMVFFVSVV IYKHFQIPCP  
 251 LPVLDSVGH LSFNNTLPMH VVMIPNNSES SDVNFMDYT HRNPAGLDEN  
 301 QAKGSLHDSH VEYEAHSDDK CEPHYFVENS FTAYAIPIV FAFVCHPEVL  
 351 ELYSELEFDRH RKMQTVSNI SITCMLVMYL IAAIFGYLTF YGEVEDELLH  
 401 EYSEVITLID PLLMVRLAVL VAVTQTVPIV LFPIETSVIT LLFPKRPFSW  
 451 IRHFLIAAVL IALNNVLVIL VPTIFYIEGF IGASSATMLI FILPAVFYLK  
 501 LVKKEIFESP QKVGALIFLV VGIFFMIGSM ALIIDWIYD PPNSKHH

# **FEATURES:**

## **Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
 N-glycosylation site

Number of matches: 5

1 83-86 NLSN  
 2 260-263 NLSF  
 3 264-267 NNTH  
 4 276-279 NNSE  
 5 369-372 NISI

[2] PDOC00004 PS00004 CAMP\_PHOSPHO\_SITE  
 cAMP- and cGMP-dependent protein kinase phosphorylation site

100 500 KNET

[3] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
 Protein kinase C phosphorylation site

Number of matches: 7

1 33-35 SEK  
 2 49-51 SQK  
 3 129-131 TAK  
 4 290-292 THR  
 5 360-362 SRR  
 6 473-475 TIK  
 7 506-508 TFR

[4] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
 Casein kinase II phosphorylation site

Number of matches: 5

1 18-21 SSGE  
 2 22-25 SAPD  
 3 129-132 TAKE  
 4 305-308 SLHD  
 5 309-312 SGVE

[5] PDOC00008 PS00008 MYPISTYL  
N-myristoylation site

Number of matches: 6

1	95-100	GLSYAM
2	153-158	SAFVSI
3	164-169	GAMSSY
4	185-191	GLEENT
5	295-301	GLDENQ
6	482-487	GASSAT

[6] PDOC00009 PS00009 AMIDATION  
Amidation site

58-61 LGKK

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	79	99	1.125	Certain
2	102	122	2.503	Certain
3	153	173	1.197	Certain
4	197	217	1.785	Certain
5	222	242	2.123	Certain
6	332	352	1.240	Certain
7	370	390	2.166	Certain
8	414	434	1.301	Certain
9	453	473	1.520	Certain
10	476	496	2.166	Certain
11	515	535	2.628	Certain

# BLAST Alignment to Top Hit:

>CRA1145000039337444 /altid=gi12017311  
 /def=gb|AAG45335.1|AF295535\_1 (AF295535) amino acid  
 transport system A3 [Rattus norvegicus] /org=Rattus  
 norvegicus /taxon=10115 /dataset=nraa /length=547  
 Length = 547

Score = 975 bits (2492), Expect = 0.0  
 Identities = 473/547 (87%), Positives = 508/547 (93%)

Query: 1 MDPMELEFVNIIEPDDESSSGESAPDSYIRIGNSEKAAMSSQFANEETESQFELTNGFLGK 60  
 MDP+ELF+VNIIEP ++S S +S Y +GNSEK AM SQFANE E SQFELTNGFLGK  
 Sbjct: 1 MDPLELFVNIIEPYEDSSSVESIQSCYTGMGNSEKGMDSQFANEAEBSQFELTNGFLGK 60

Query: 61 KKLADYADEHHPGTTSEFM:SENLNAIMGSGILGLSYAMAYTSVILFIIMLLAVAILSL 120  
 K L DYADEHHPGTTSEFM:SENLNAIMGSGILGLSYAMA TG++LF+IMLL VAILSL  
 Sbjct: 61 KTLTDYADEHHPGTTSEFM:SENLNAIMGSGILGLSYAMANTGIVLFVIMLLAVAILSL 120

Query: 131 YSVHLLKTAKEGSSLIYFHLGEFAFGWPGKIGAFVSITMQNIGAMSSYLFIIKYELEPV 180  
 YSVHLLKTAKEGSSLIYFHLGEFAFGWPGKIGAF+ SITMQNIGAMSSYLFIIKYELEPV  
 Sbjct: 131 YSVHLLKTAKEGSSLIYFHLGEKAFGWPGKIGAFISITMQNIGAMSSYLFIIKYELEPV 180

Query: 181 IPAFMGLEENTGEWYLNBNYLIIFVSVGIILPLSLLKNLGYLGYTSGFSLTCMVFFVSVV 240  
 IE FMGLEENTGEWYLNBNYLI++FVSVGIILPLSLLKNLGYLGYTSGFSLTCMVFFVSVV  
 Sbjct: 181 IRVFMGLEENTGEWYLNBNYLVLFVSVGIILPLSLLKNLGYLGYTSGFSLTCMVFFVSVV 240

Query: 241 IYKKFQIPCPPLVLDHSHVNLSEFNNTLPMHVVMLPNNSESSDVNEMMDYTHFNFAGLDEN 300  
 IYKKFQIPCPPLVLDH+ GNL+FNNTLPMHV+MLPNNSES+ +NFM+DYTHF+F GLDE  
 Sbjct: 241 IYKKFQIPCPPLVLDHSHVNLSEFNNTLPMHVVMLPNNSESSDVNEMMDYTHFNFAGLDEK 300

Query: 301 QAKGSLHDHSGVEYEAHSDHKCEPKYFVFNSTAYAIPILVFAFVCHPEVLPIYSELKDRS 360  
 A G LH SGVEYEAHS HKC+PKYFVFNSTAYAIPI L FAFVCHPEVLPIYSELKDRS  
 Sbjct: 301 PAAGPLHGSGVEYEAHSGDKCQPKYFVFNSTAYAIPI LAFAFVCHPEVLPIYSELKDRS 360

Query: 361 RRMQTVSNISITGMLVMYLLAALFGYLTIFYGEVEDELLHAYSFVYTLDIPLLMVRLAVL 420  
 RRMQTVSNISITGMLVMYLLAALFGYL+FYGEVEDELLHAYSFVYTL D LLMVRLAVL  
 Sbjct: 361 RRMQTVSNISITGMLVMYLLAALFGYLSFYGEVEDELLHAYSFVYTFETALLMVRLAVL 420

Query: 421 VAVTQTVPIVLFPIPTSVITLLFPKRPFSWIRHFLIAAVLIALNNVLVILVPTIKYIFGF 480  
 VAVT TVPIVLFPIPTSVITLLFP+RPFSW++HF IAA++IALNNVLVILVPTIKYIFGF  
 Sbjct: 421 VAVTLTVPIVLFPIPTSVITLLFPKRPFSWVKHFGIAAIIIALNNVLVILVPTIKYIFGF 480

Query: 481 IGASSATMLIFILPAVFYFLVKKETFRSPQKVGALIFLVVGIFFMIGSMALIIDIWIYD 540  
 IGASSATMLIFILPA FYFLVKK E RSPQK+GAL+FLV GI FM+GSMALIIDIWIY+  
 Sbjct: 481 IGASSATMLIFILPAAFYFLVKKELRSPQKIGALVFLVTGIIIFMIGSMALIIDIWIYN 540

Query: 541 PPN SKHH 547  
 PPN HH  
 Sbjct: 541 PPNPDHH 547

>CRA1114000033649823 /altid=gi10945621  
 /def=gb|AAG24618.1|AF298897\_1 (AF298897) amino acid  
 transporter system A [Homo sapiens] /org=Homo sapiens  
 /taxon=9606 /dataset=nraa /length=506  
 Length = 506

Score = 597 bits (1522), Expect = e-169  
 Identities = 315/549 (57%), Positives = 383/549 (69%), Gaps = 46/549 (8%)

Query: 1 MDPMELRNVNIIEPDDESSSGESAPD---SYIRIGNSEKAAMSSQFANEETESQKFLTNGF 57  
 M E+ +I PD++SSS S D SY +++AA+ S +A+ D E+Q FL  
 Sbjct: 1 MKKAEMGRFSISPDEDSSSYSSNSDFNYSY----PTKQAALKSHYADVDPENQNFELLESN 56

Query: 58 LGKKKLADYADEHHPCGTTSPGMSSEFNLSNAIMSGSILGLSYAMAYTGVILFIIMLLAVAI 117  
 LGKKK Y E HPCTTSPGMS FNLSNAI+GGGILGLSYAMA TG+ LFII+L V+I  
 Sbjct: 57 LGKKK---YETEFHPGTTSPGMSSEFNLSNAIVGSGILGLSYAMANTGIALFIILTFVSI 113

Query: 119 LSLYSVHLLKLTAKEGGSLIYENLGEKAFGWPGKIGAFVSITMQNIGAMSSSYLFIKVEL 177  
 SLYSVHLLKTA EGGSL+YE+LG KAFG GK+ A SITMQNIGAMSSSYLFI+KVEL  
 Sbjct: 114 FSLYSVHLLKLTANEGGSLIYEQLGKAFGLVGKLAASGSITMQNIGAMSSSYLFIVKVEL 173

Query: 178 PEVIPAFMGLEENTGEWYLNENYLIIFVSVGIILPLSLLKNLGYLGYTSGFSLTGCMVFFV 237  
 P VI+A +E+ TG WYLNENY++ VS+ +ILPLSL +NLGYLGYTSG SL CMVFF+  
 Sbjct: 174 PLVIQALTNIEDKTGLWYLNENYLVLLVSLVILPLSLFRNLGYLGYTSGLSLLCMVFFL 233

Query: 238 SVVIYKKFQIPCPPLVLDHSAHLSFNNTLPMHVVMLPNNSESSIVNFMDYTHRNPAGL 297  
 VVI KKFQ+PCP- + H + N TL ++P  
 Sbjct: 234 IVVICRKFQVPCPVEAA--LIHETINTTLTQPTALVP----- 269

Query: 298 DENQAFGSLHDSGVEYEAHSDFCEPKYFVFNSTAYAIPILVFAFVCHPEVLPIYSELK 357  
 + + +D C P YF+FNS+T YA+PIL+F+VCHP VLPYI ELK  
 Sbjct: 270 -----ALSHNVTENDCORPHYFIFNSQTVYAVPILIFSVCHPAVLPIYEELK 317

Query: 358 DRSRPFMQTVSNISITGMLVMYLLAALFGYLTFYGEVEDELLHAYSKVYTLDIPLLMVRL 417  
 DRSRP+M VS IS M +MYLLAALFGYLTFY VE ELLH YS + DI LL+VRL  
 Sbjct: 318 DRSRPFMMNVSKISFFAMFLMYLLAALFGYLTFYEHVESELLHTYSSILGTDILLIVRL 377

Query: 418 AVLVAVTQTVPVIVLFPFIFTSVITLLFPKPFPSWIRHFLIAAVLIAALNNVLVILVPTIKYI 477  
 AVL+AVT TVP+V+FPFIF+SV LL + FSW RH LI ++A N+LVI VPTI+ I  
 Sbjct: 378 AVLMAVTLTVPVVIFPIFSSVTHLLCASKDFSWWRHSLITVSILAFTNLLVIFVPTIRDI 437

Query: 478 FGFIGASSATMLIFILPAVEYELVKKETFRSPQKVGALIFLVVGIFFMIGSMALIIDW 537  
 FGFIGAS+A+MLIFILP+ FY+FLVKKKE +S QK+GAL FL+ G+ M GSMALI++DW  
 Sbjct: 438 FGFIGASAASMLIFILPSAFYIKLVKKKEPMKSVQKIGALFFLLSGVLVMTGSMALIVLDW 497

Query: 538 IYDPPNSKH 546  
 +++ P H  
 Sbjct: 498 VHNAPGGGH 506

# Hammer search results (Pfam):

Model	Description	Score	E-value	N
PF01490	Transmembrane amino acid transporter protein	187.0	2.9e-52	2
CE00398	E00398 CD53	4.0	4.8	1

## Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00398	1/1	90	110	..	1 23	[. 4.0	4.8
PF01490	1/2	99	236	..	1 179	[. 58.9	2.5e-14
PF01490	2/2	305	529	..	200 467	.] 133.9	3e-36









9451 TTTTGTAAATG TGCTTACCTC TAAGACAAAA AAACCTGGCT TTCCTTATTA  
 9501 ATTATACATA CATTAAAAAT GAATTAGGAA GTTAAGATC ACTGATGAAT  
 9551 AGAAATAGGA AAAACTTCCC CCAATCCCAC AGTATAGAT CATCTTCATG  
 9601 AGAGAAGAAAT GTTCCACTTT TFAAAATGAG GGCCTCATTT TAGGCTTATA  
 9651 AACACTTAGC AGATGAATTT GGTGAGAACA ATTAATCAC TAAACATCAT  
 9701 GGGGTGTGTT TTCTGTGTCT AAGTAGCCCA GACTGATTA AGCTTTCTCT  
 9751 CTAAATTTAT AGCAAGTGAC ACAGTATTTT AAAATTTTAA CTCTTAGTAT  
 9801 TTTCTGCCAG AGAAAGTACA TGTTTAGAAT ACAGGGAATG CTATTTATTT  
 9851 TTCCAGGGAA CAAATTTATA TAATCTGAAT TACATTTTTC CTAAAAACA  
 9901 GTTAAGTTCA TAAGGATAT GGAATAATAT AGGAATAAGT CATGCTTAG  
 9951 ACAGTTCTTG CAAATTAATT CTATGGAAAA TAAGAGTACA ACATAGCTAC  
 10001 AGGGCTTATA AAATTTATAA TTCATGGTCC AAATGTACAT TTCTAGTATT  
 10051 GATTTTCATTG GAATTTACCA AGGGATTAGA TCAATCTGCG GGAAGTGTAA  
 10101 TTTTTTAAAA ATAAACAAAG ATAAAGATTT TTTTCTGAAA TTCTAGGTAA  
 10151 AAGGCAGCAT TGGCTTCCCA TTTATTACGT AGATGCTTCT ATCAACATTC  
 10201 TTATTTTGTG GTTCCAAATC TTGGATTTGG AAAAAATACA ATGGATATA  
 10251 ACATAAAGAA ACATACATG CATGTGGGGA TCTAAACAC AGAATGACAT  
 10301 CTGAATGTAA AAAAAAAAAA AAAAAAAAAA GGAATTTTTC CTGCCCCATC  
 10351 CTTAGCTTTG TGTGCTTTCT CTATTATATA TSCAACTGCT TGGCTTCTTA  
 10401 TCTTACAAAG TACTTGTAA TCTAATGCAC AGGATCAGCA GTAAATGAGC  
 10451 TCAGACTGCA TGGTTTGGCC TTGCGATTCC TAGATTTTAC ATTAAGTTTCT  
 10501 AGTCAGGCTA TTGAATAGCC CTTCATTTCT AAGTGTCTAT GTTAATATCA  
 10551 TSCAAATATG ACTTATATAT TCCCATGTGC TGAGTAAATA GATGTAGCAT  
 10601 TTGCTAATGT TGGTACATAT TTAGCATCTA AGTTATGAAC CAGATTTCTAC  
 10651 CACTGGGTAA CATTAATAAA AAGTTAGGGA CTTCAGGTAT GTAAATATA  
 10701 GCAAAATCTA TTTCTACGAC TTTAAAGGGT ATGTGTAGAG TTCTGAAAAAG  
 10751 AATTTCTCAG GTTCCGCCAA ATCCACATAC TTTTGGAAAG CTGATGATTG  
 10801 AAAAGATTA TGTGATCTTT TATTGTAACT TCTAACATAA TTACATTTTA  
 10851 TTTATTGTAG AAATTTTATT AACTACTCTA TCTTCCCTTC GCAATATCAT  
 10901 GCTGCTTGGT GTGGCAATAT TATCACTGTA TTCAGTTTCA CTTTTATTTAA  
 10951 AAACAGCCAA GGAAGGAGGT ATGCTACCAC TTGAGTCCAA CACATTTCTAT  
 11001 TTTAATTTCT ATAAAGAGT ATTTCACTCT GTTGTCTCAT AACCTTAGGA  
 11051 TGATTATAGT CAGTTTCCACA TTTCAATTTT TCTGAGGCC AGTGACACGA  
 11101 TCTCTCAGTG TTTATAGTTG TTTGGGCAAG TGAGAGGCCAG GAGTGAAAGT  
 11151 CAAGTGGCTC AGTTTCAAGA CAAATAGAAA AAAGAAATTT CTGATATATG  
 11201 ATAGAAATTA CTGTTTGGAC TTGCTACATG CAGCTAAAAA AAATAAAACC  
 11251 ATTGATTCTT GTTGGAGAAA CATTTTGATA TATTGCTTAT TGGTTTGTGA  
 11301 GGTGTGCTCT TTTGGGCTTA TAATTTCTAT ATGATGTTTA TTTACATGTT  
 11351 TGAGACTCCA GATGGAATT ATATGACAAA AATATTTTAG TCATTAAAAAC  
 11401 AATCTCTTTA ACAAGGCTAT TTTATCTTTG ATTGTAGGGT CTTTGATTTA  
 11451 TGAAAAATTA GGAAGAAAGG CATTGTGATG GCGGGGAAAA ATTGGAGCTT  
 11501 TTGTTTCCAT TACAATGCAG AACATTGGAG GTAAGGGGAT ATACTTTCCA  
 11551 ATGGATCCCA TAAACTTTCT ATAGCGTGTT CAATAAAATA GAAACTTTAT  
 11601 GGCAATAAAC AGGCACTTTA GATACAGAAA AATTGCTACT TATAGTTCTT  
 11651 AAATTTTAAA ATGATAGTTT CTAAATAGG TTTGTGCTCT GTTTTAAATTA  
 11701 AAAACAGCAA TATCTAAGAA TGAAATAACA TATAAAACCC TGCCAATTGA  
 11751 ATTCTAGAAT TAAATATATA AATAAAAGCT TTCTTGATTT TTAATGTTAT  
 11801 TATAGCATGA ATTATTACTC TAAAAAATG AAGAATTTGT GTTTATATCT  
 11851 GTCATTGACA AAACAGTTGA CGTTTTCTAT GTGTGATGTA GTTCGATTTA  
 11901 CTAAACTGAA AAGTGGGTGT CTGGGGGAAC ATAGCCAAAT GTGTGGTCC  
 11951 TTGAACGCGA GCTGCACTG AGCCAGCCCA CTAGACAGTG TCTCTGGAAG  
 12001 TTTACTAAGG CAAAAGTCTG GCTAGGCATC AAATGCACTA TAAACCCCGG  
 12051 TTTGTTGATT CTATGGATTC TTATAATTCC CACTGAATTA TCATTGCCAG  
 12101 TGTAGGACCT AGAAATATAT ATATATATTT TTAACAATGT TCTCTGTTG  
 12151 GTGTGTTTGC CCAAGGCTTT CATACTGTTT CTGTTGTGTC TTTGGCCCTC  
 12201 AGAAGGCATC CAAAGCCATA TTTCAGATGT CCTGCGGCTT GTTCTGTGGC  
 12251 ACATGGCCCC AGGCACTCTC CCACATAATG ACATTTACTC CTTCACTCTC  
 12301 TACCAGTCC CTAACCTGCG TATTTCTATT CTCTGATCTT TCTTTTCTCA  
 12351 GTGAATACCA CAGCACTGCA TCCAGTTTCT GAGGGCAGAA ATCTGATATG  
 12401 CAGCGTAAAT GTTCTCTTTT CCCCCAATCT GCATGTCCAA TCAAAATGGCA  
 12451 AAGTCTGTTT ATTTGATCTC TTACTTATCT CTTGAAACCT TCTCTCTGTG  
 12501 CCGTCTCTAT GACCAAGAT GATCACCATT TATAGCTCAG ACTATTGCGAG  
 12551 TAGCTTCTA ACTGCTCTTC CTGGCTTGAG TTTCCCTTSC TCTCAGATAA

FIGURE 3, page 4 of 23

13651 13701 13751 13801 13851 13901 13951 14001 14051 14101 14151 14201 14251 14301 14351 14401 14451 14501 14551 14601 14651 14701 14751 14801 14851 14901 14951 15001 15051 15101 15151 15201 15251 15301 15351 15401 15451 15501 15551 15601 15651 15701

13601 ACTCTAATTT GTTCTCCAGA TAAACTTTCT CAAATTTGAG TGTGTTTCTA  
13651 CTTTTGTGCT GCATAAAATT CTTGAGCAAG CTTTATTAT TTTCAAGSAA  
13701 AAACCTAAAC TCATTGCACT GACACAAAGT CTTGCTCTAG TTTCTCTGCT  
13751 CAATCTTTCT AAACCTTTCT AGCAATGCCC ATATCTATCT ATCTTTTATCT  
13801 ATCTATCTAT CTATCTAT T ATCTATCTAT CTATCTATCT ATCTATCTAT  
13851 AATTTATCCA TCATCTATAC CTTACATGTC CTGCTGCAAA CCATAA-CAAA  
13901 TTATATTTAT TCCCTTAACA GTACTATTTT AATATTTTAA AAAATCATTC  
13951 ATGCTTCTCT TTCACAGGCT ACTTTCCTCC CTGACTGTCT TGTCAAAGTC  
14001 CTCCTACCCCT AACACACACG CACACACACA CACACACACA CACACACACA  
14051 CACACACATT TTCTCTCTCA CTCTGCTCAC CTGCTCTATT GCTCTCTTAG  
14101 ACTGGTAAAT ACTAGTTCTT CTGCTCTCTC ATGCTCTCTT TTGATCTTAG  
14151 TATGTTACTG TTTTCTAAAG GATATTTTAA AACACTTGAG TAGACAAATA  
14201 GCTTTTGGAG TCTGATGGAC CTCAATTTGA GTCTGTTTCT GTCACTATCT  
14251 GTGAACCTGG GAAGATCACT GTACTCTCTT GTCTGATTTT TGCATGTAATA  
14301 AAAATTACCT TACAAAGGCT ATTGTSAGGA TGAATAAGG TAACATATGG  
14351 CACATAATAA GTGTCTGTA TATGTTCTCT TCTCTCTGCT TTCTCTGCTT  
14401 CCATATCCAT GTCTCTGGAG TTCTCTCAAT TATTTTAA ATAGGCATTT  
14451 AAAAAATTAT AAAAAATAA TATGATGATT GTGAAAACT AAAACACTGC  
14501 ATAAATATAT AAATTACCAA GAAAAATTTA TGTCACTCAT CTTCAAGAAAT  
14551 AACTACTCAT AGGTTTCTCC CTATGCTTAA TTCAACAAAT ACATTGAATA  
14601 TTGTTAGTAT TGGATCATCT TATGATACCG ATTTTCAGCT TTCTTTTAA  
14651 ATTTAACAAT ATGCTTGA AATATTTGCA TGTATTTCTT TTTAATGATT  
14701 TTTGAGGTTT CCATTACACA AATGTGCTAT AATTTGTTA CAGTATCTCT  
14751 ATTGATGAAC AGTTGGATTG TTTCTAATTT TTTCACTGTTA TAAAAATGCT  
14801 ACAGTAAATA CACTTGCACA GAGATCTTGC AAACAGGCAA CCACTTTTAA  
14851 TAAATAAATT CACTGGAGTT ATCAAGGATT TCTGGAATGC AGAAATTTCT  
14901 TTAGTAATCT ATCTAACTAT ACTCACCTCG ATAATGGATA GTTGGTAAGC  
14951 AGATAAGTAA AATTCAAGCA TATCTTATGA TTGTGTTAA AAAAAATTTT  
15001 ATATGTTAAG ACTACATCT TGGTAGAAT TTGACAGTAA TATCAAAAT  
15051 GTCTCTACTT TTTTACTGGT TTGGAGCCAT ATGCATATTA GGGCCCCAAA  
15101 TCCCAACAAA TAGACCACTT TACATTTGTT TCAAACTCTC AGCCTTATCA  
15151 AGGTTTAAAG TATCGAGCAT TTCTATAGGAT TGCCTTATAG TTGCTCTAAT  
15201 TTAACAACCTG AAATAACCAG GCATAAGCAT AATTAACCTT GGACTCAAGA  
15251 AGTTGAGTGG CAGCACCTCA GCTGTGTTTC AAAGCATAGC CACTACTACG  
15301 CTTCTAAACA ATGGAATAAA GTATAAGCG GTCTCTCAGT CAAGCCTCAC  
15351 ACAGGTAAAG GCGGTGACTT TAAGGGAGTA AGATGAAATA TCGTAACATC  
15401 ACCCCAGAAA TAATGCTCTC ACTTTGGTTA CTTTATTGTA TTAGTTGATA  
15451 TTTGGCATAA GAGAAATCAC TTGTATTTCT CTATTTAACA ACTCTACATT  
15501 TAGAACACTT AATTTTCTCA ATCCCCTAAA AAATTAACAT TTAGTGCAGA  
15551 TGTTTTTACA TTAACAGATT AATGTCTGGA TCATTCTGAA TTTTGAAGA  
15601 CCAAACATGT TAACATCACT GACATCACTG AAAACCAGCA ATTAATAGCT  
15651 GTAACATTGA ATGGTACCTC ACCAAGCCAG CTAATCAGAA ATATCTCCTG  
15701 TGTTCACACT CTGTAAGATT TAGCTTTAGC CAAGGTCTTT GCAAAGATTA  
15751 ACCAAATAAT GTGTACAGAA GGTACATCCG CTATTGTAAA AATCATTTCA  
15801 CTTTGACAGT ACAGAAGAG CACGAGCCT TCTGTTTAG ATGTAGTCCG  
15851 TCCCTTTCAA GCTGTATGAT TGTGGACATG TCAACTTAAC ATCTCGGAGT  
15901 TTTTATATCT TCATCAGTGG AATGAGAATA ACAACATATA TCTTGTCTATC  
15951 TCACAGGGTT TTTTATATGA TCAAAATGAAG TAATGTGCAG AACTAACCAA  
16001 TGTGGGGAAT TATTATCATC ACTGTTACTT TCATATGAAG TGAAGAAAAAT  
16051 ATTTTAAAC TCAGTAGTTT AATTTACAAT TTAAGTATGT GTTTTAAAGT  
16101 GCCTGTTAGC AAAAATTCAG TAGAAGGATG TAGGACACAC TTAAAGTTTT  
16151 CATGTAAAT TTGTGAGTTC TATTTTAAAC TGAATCTTTT GGCCATGTGT  
16201 CAACAAATTA ACGTTATCCT TCACCAATAG GGTGGGCTTG AAAAAGGGGT  
16251 GATGCATAAA TATTTACAGT TGTAGGCAAA ATTGTAATGT TATGTATATG  
16301 AATACATATT CATTTTTTCA GGGAGAAGGC TTGTAGATTT CATCAAGAAA  
16351 TCTTTCACAA GAGTAGATAA TCATTCATGT ATCACTTACC TAGATGCTCA  
16401 TGAATTTTGG CCACTTTTATA TAATTCCTTA GTTAGCCAAA AGGAGAGTAA  
16451 GATGAAGAGG GGGGAAAAAA AAAACTTCTT TGACAAAGAT GGAGAGAAGC  
16501 TGTCTCTCTT TGTATCTTT TATCAATCCA GGAAGCCTTT GGTTTTGACA  
16551 ATAAAGTGGT TGAGACTTTG TGTACTCCTC AGATAGGTCC CGGAGGACTA  
16601 GATTGGTGCC CATCTGCAGA AAAACAGAGG GGATATATTG ACTCTGCAGA  
16651 TCTGCCCTTT GATTCTGCCA TCTCTCAGCT GGCCCATGCC TTTTGTGGC  
16701 AGACTACTGC CCAAGTTATA GACACTAACA CAGGCACACT GAGTATGGGC

FIGURE 3, page 5 of 23

15751 TATTTTGATT TATAACTAAT GAGGGCAGAA CCTTAGAACT GCAGCTTAC  
15801 TGTAACCTTT GGAGCAGGAT TTAACACAGA ATCAGCCCTG ATAGTGTAA  
15851 CAAAGGTCCA CCTGAAAGAG CTGAAAGGTC AAATGTCTAT CTTGGAAGAG  
15901 AACTTGGGAG CAGTGGCAAA TACACAAATGA CTTTTTTTTT CATTPSGGGG  
15951 AATTAGATGT CATCTTACAT ATCCCAAAATG TCATAACTTG CTTGCAAGGG  
16001 ACTTCAGTAC TGTCACACG ATTAAGCTGT CACATTTTCC ATTTTASAA  
16051 TGTCAAGCTA CCTTTTTATC ATTAATATG AACTACCTGA AGTAATCAGA  
16101 GATTTCATGG GACTTGAAGA AAATACTGGG TATGTCTTAT GCTCCTTG  
16151 TGACATCAAG TGAATCATTC TACTTGGTCT TTTCTGATTC TAATATCCT  
16201 GTCTCTCACT TGTAGAGAAT GGTACCTCAA TGGCAACTAC CTCATCCTAT  
16251 TGTGTCTGT TGAATATTAT CTTCACCTTT CGCTCCTTAA AAATTTAGGT  
16301 AAGATATTT TCTAACTGGA AATATTTTTA TTTTATTTT ACATTTAAAT  
16351 AGTTAGCTA ATTGTAGATG CCATATTCAC CTTCCAAAAT GCTTCTTCTA  
16401 ACTTCTAGGT TATCTTGGCT ATACAGTGGG ATTTTCTCTT ACCTGCATGG  
16451 TGTCTTTTGT TAGTGTGGTA AGTGTGTGA TGACATGATC CTTGCAAGGT  
16501 GGTAGCATG AGTTTTTTT TGGCTAAAT AGTGTCTCA TTTGTCTTAA  
16551 GCACTTCACT AATATGAAAT AGTCTTGTGA TCACAAGTGA TTTTCTTCTA  
16601 GACTAATTTA GAGCAAAAAA AGAGCAGCTA CGATTTAAAG ATAGTGGAGG  
16651 TAGAATATCA AAGCTACTAC TAATGGTTTG GTCTAGGCAC ACTGTTTATA  
16701 TATGGGGAAA AAAGGAAAAA TCCAGCAGG AACATGACAA GGCTATATCC  
16751 TTTAGAACAG CAGAGGAGAG TCCAGATGA GAAACAAGAA GGCTATATCC  
16801 ATATTACAT GAATCAGCCA TTCTCTCTTA CACATTCAC CCATTAAGAG  
16851 AGGACAAGAA CAGTGGGAT AAAGAAGAAA TCCTCCTCTC TAGGTCCTCTG  
16901 ACAAAGAGG GAATTTCTTG CACTATCATG AATGCCAAAA TTTATAAAGG  
16951 ATTTCCCCAA ACAGGTAAAG GAGAAGGAAA AAAAGTTTGG AAGACCCAGG  
17001 TCACCTTAGT TTGAAGAAAT AAGGAAATGA TCATCTTTCT CATGGAAGGG  
17051 CATGAAAGAG GGTGGGAAGG ATTCTTGCAA AATATTGTCC TGTTAACTCT  
17101 AAGAGGCAGG GCTGCCAATC ACAGCTCCAA CTCTCCCTT AGAACACAGG  
17151 CTAGACCAAC TTTACTTTCT CCACTACTCT AAAAGGAATC CTAATCTAG  
17201 TTCCCTCACC CCCACCCTA TAAGCCACAC ATATGGATTC TTATTTCAAT  
17251 GTTTTTTCTC AAAAAGCTGA TTTTTTTTTT TTTTAAATG ACTGAGTCTA  
17301 GGTGATTAC AAGAAATTC AAATACCCTG CCTCTACCT GTTTTGGATC  
17351 ACAGTGTGG AAATCTGTCA TTCAACAACA CGCTTCCAAT GCATGTGGTA  
17401 ATGTTACCCA ACAACTCTGA GAGTCTGAT GTGAACCTCA TGATGGATTA  
17451 CACCCACCGC AATCCTGCAG GGCTGGATGA GAACCAGGCC AAGGGCTCTC  
17501 TTCAATGACAG TGGAGTAGAA TATGAAGCTC ATAGTGATGA CAAGTGTA  
17551 CCCAAATACT TTGTATTCAA CTCCCGGTA AGTGAGCGGT CCGGGCTTCT  
17601 AATGACTACA GTTATGTGTT TTCTAAGTTT TTATTCAATA AACTGACATG  
17651 GCCTGAGATC ACCATCTATG TTGGAATGCT AAACACGTGG TGTGTCTTT  
17701 GTTTTTCAGA CGGCCTATGC AATTCCTATC CTAGTATTTG CTTTGTATG  
17751 CCACCCTGAG GTCCTTCCCA TCTACAGTGA ACTTAAAGAG TAAGGCAGCC  
17801 ATCATTTTAG CATTCTAATT TGCTTTGAAA TTCTGCTCAT ATGTTCAAAG  
17851 ATTCTTTAAC AGGAAACACA GTTATAGCT TCCTCTCAG AGAAAAATG  
17901 TACTCCATCC ACTCCTCAGT AACATGCTTT AATCAGAAAG GTGGGAATCA  
17951 GCCCACCACA GCACTACCTT ATCTTCTTTC TCTCCTTCT CTCCACCATA  
18001 ATGGTTACAG GGAGGGGTTG ATGGCAGGTG GACAAGGAGT CGATGGTTGT  
18051 AATAATTTTG GCAGGTGTTG GGAATTTAAA TTTGAATTTT GTTCGGAAGA  
18101 AATGATGTCA GCTGGACTAG AAATGAAAAC ACCCATGACG ACCAAAATTT  
18151 ATGGTTAGGG GCAGCCTCGA TAAGCCAGTG ATGTCATTTA TAGTCAGCAC  
18201 CTAACCTTG TCTAGAACA ATTCAATACA AGAGATGTGT CAATATCTGT  
18251 CCTTGTGTT CTTATTTGTA CAATAGAGTC ACTGGCTAGA AAATCTTGT  
18301 TCTTCCAGCT GATGGTCTAT GGTTCAATTG TATTCTTTT CTTTGAAGT  
18351 TGTGATATT TGCTTGGGAA CAAAGGATAT GAACCTATTA TAGCTTTTT  
18401 CCTCTTCTCT TTAAGGGAGG ATATTATATA ATAATTCTCA ACTTCTTTAA  
18451 TCTAGACATC AGTAACCTCA GTCTTCATTC TCACTAAATA GCAAACTTT  
18501 CCCCATAAAT TCTGATTTAC CTCATAAAAA ATTTGAGAAC ACTTTCAAGT  
18551 ATTTTGATGT CTTGATTTA CTTTGAAGAT TACATGTAGC AGTTACTCCA  
18601 GAAGCCTGAC AATTGATCTT TGGCAGCCAG GTTCCTTCTA GAATGTTTT  
18651 CAGAAGCTTT TCAGGTAGTC TGGACTCTG GCAGTAGTAC TTTGCTGACT  
18701 CTACTAGGTT CTTTCTCTCA TTTAAAGTCA TCTCATTATG AAATGCAAAA  
18751 GCTTCTATG TTAGGAGCT GTTTCATCTT TATGTTAATT ATATTCTTAT  
18801 TCAGTGGGCA AGCTTACTGA CCTACGTGAA ATAGACTGTT CCTCTCTAG  
18851 GGAATGATT GTTTTAAAG CTGAAGGACT AGTGTTTAAG AAAAATGSA

FIGURE 3, page 6 of 23



[illegible]







31501 GATATTTAAT CCTAAATTAT TATGATGATT ATAATTTTGG CATCAGATAT  
 31551 ATACCACCTA SAATGAATGT GGAAGAAATG ASTCTTTTAT GGTTAATTTG  
 31601 AAAGAATCCA TTGAAGATAG AAAATGAGAG AATAGAAGAA ACCTGAGAAT  
 31651 AGTAAAATAA AGAGCAGAGA AAATATGGGG GCAGGGAAAA CATGTAGTGT  
 31701 CTAAGGATTG ATTATGAATG AACGATTAGG GGGATTGATG GATCAGAGGG  
 31751 TAAGTATATG CTAACTTTA TAAGAAACTT CCACATAGTT TTCCACAGTG  
 31801 TTTCTACCAT TTTCATTTCC ACCCGTACTA CCTACAACCT CCCTGACTC  
 31851 CACAGCCCTG CCAACATTTG GTGTTGTCTT TTGCATTTTA GCTTTTCTAG  
 31901 TGGTCTGAA ATGGTAACTC ATTGTGATTT TCATTTCTGC TTCTGTGACA  
 31951 ACTAATGTTG AAAACTTTTC AAGTGTTTAA TGGTCACTCA TATATCTTCT  
 32001 TTTGTGAAGT GTGTATTCAA ATCTTTTGCC CATTTTTTAA ATTTAGGTTA  
 32051 TGTGTTTTTA TTGGGTATTT GTAGAAGCTC TTTAAATATG GATCCATGTC  
 32101 CAGATTGCCA ATATATTTTC CCAGTCTATG GTATGGTTGC TTATTTTCTT  
 32151 AAAGGTGTCT TAATTACATC TTTCTGGGGC CAGGTCACCA TAGCTCAAAG  
 32201 TTTTGCAATT TATGTCTTAA TGAGATAATA TTAATCAGAG TGGTATAGTC  
 32251 AAAATTAAAT GTTTTGATGT CCTGGGCCCA TATAGGTAGG ACTGGATCAT  
 32301 CTAACCAAGA TGCACAAAAA AAAAAACAAA AAAACAAAAA TAGTACTTGG  
 32351 AAAAATTAT TTTAAATTAA ACA

# FEATURES:

Start: 3000  
 Exon: 3000-3118  
 Intron: 3119-7452  
 Exon: 7453-7543  
 Intron: 7544-8039  
 Exon: 8040-8155  
 Intron: 8156-10844  
 Exon: 10845-10968  
 Intron: 10969-11127  
 Exon: 11128-11530  
 Intron: 11531-16047  
 Exon: 16048-16129  
 Intron: 16130-16215  
 Exon: 16216-16298  
 Intron: 16299-16408  
 Exon: 16409-16467  
 Intron: 16468-17301  
 Exon: 17302-17577  
 Intron: 17578-17799  
 Exon: 17710-17789  
 Intron: 17790-19073  
 Exon: 19074-19174  
 Intron: 19175-20904  
 Exon: 20905-21029  
 Intron: 21030-26649  
 Exon: 26650-26794  
 Intron: 26795-27670  
 Exon: 27671-27768  
 Intron: 27769-29273  
 Exon: 29274-29372  
 Stop: 29373

# CHROMOSOME MAP POSITION:

Chromosome 12

# ALLELIC VARIANTS (SNPs):

DNA				Protein
Position	Major	Minor	Domain	Position Major Minor
1386	T	C	Beyond ORF(5')	
2594	T	C	Beyond ORF(5')	
2757	G	T	Beyond ORF(5')	
6107	C	T	Intron	
6392	T	C	Intron	



ATTTTTTTCTTTTTTTTAA GAATTCCTTTTTTTGACTCTTTTTGCAACAGECTGTTCTCC  
TTTTATATTCCTTTATAATGTTTTATTCTGTSAAGSTTATTCTCTATTCTGAATCTTT  
TCTTTCAAAATGCTTTCTTTTTATTAAATTAAATGTAAGTCCCTTTTAAATGCTTTG

2594 CTGAACCTTCTTTTTGTTACTATTCTTAACTTTGGCTTCAGGATCCAAGTGCCCTA-AAAAGT  
TACTTCTAAACTTGATCTCACCCTATGTTGCATATTATCAAGCATTGCTGGGTCTAAT  
TCTTTCTATGTTCAATTAATTAAGCAGTAATTTCTTTCTAGTTATTGCTAGTAGAGAC  
ACTGGTAGATCTGCCTTGGTAGACCTTCTCTGTGCAACAATTTACTTTGTCTCTCTTT  
CTTTTAAAAATATGTATCCACTCACAATACTTAAATTTCTTTGAAGACTGCTGCCATG  
[T,C]  
TTAAGATTTCTTTTTTTCCATAGTGAATAAATAAAGCTGCCATTTTCATTATACATAG  
GCACTCTATAAATATCTGCTAATTTAGCAATTATTAGTAATTTCTTTCTCTCTCTCAT  
TCTCTCTTTCTCTGATTGGGTAAGGAACATTCAGGATTTGGCTATGTAAGSTTTTCA  
GGAATTTCTTTCTCTCTCTTTTACAGAGAGCATA-AAAATGTAGATGATTCTATTTCT  
ACTTATTTCTATTTAAATAAAATTATAATGATGATGTTGTGTTCTGTTTGCAGAACAGAG

2757 TTAATGCTTAGTAGAGACACTGGTAGATTCTGCCTTGGTAGACCTTCTCTGTCAACAAT  
TACTTTGCTTTCTCTTTTAAACATGTATCCACTCACAATACTTAAATTTCTCT  
GAAGACTGCTCTCTATGTTTAAAGATTTCTTTTTTTTCCATAGTGAAGTAAAGCTTGC  
CATTTTCTATATACATAGGCACTCTATAAATATCTGCTAATTTAGCAATATTAGTAAT  
TCTCTTCT  
[G,T]  
CTTATGTAAGSTTTTCAGGAGTTTCTTTCTCTCTCTCTTTTACAGAGAGCATA-AAAAT  
GTAGATGATTCTATTTCACTTATTTCAATTTAAATAAAATTATAATGATGATGTTGTGTT  
CTGTTTGCAGAACAGAGTGTCTGAACATCAACACAAGTGGAGAACCTTAAAGCTGAAG  
GTACACTATATTTTACACTGAAGGGGCTTGTGTGTGGACAGAAAGCGCTGACAGCTC  
AAATGGATCTCTGGAAGTGAAGATGTCAACATCAACAGATGATGAGAGCAGCAGTGC

6107 CT  
GTCT  
GTACAGATCT  
TTGAGACAGACAGAGTTTCACTCTTGTGGCCATGCTGGAGTGCAATGGTGCAATCTCGG  
CTCACTGCAACCTCTGCCTCCAGGTTCAAGCGATTCTCTCTCTCTCTCTCTCTCTCTCTCT  
[C,T]  
TGGGATTACAGGCGCCTGCCACCACACTCGGCTAACTTTTTGTATTTCTAGTAGAGACGA  
AGTTTCAACATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAAGTTATCCACCCACCT  
CAGCCTCCCAAAGTGGTGGGATTACAGGCGTGAGCCACTGTGCTGGCCTCTAGGATTAT  
ATTAATAGAACCAATCTTCAATTATTTTATCTTTCTTTATCTTTCTTTTCTATGAGAAAT  
GTCTTAAATTTTCAAACCTCAATTTGAAAGCACTTTTAAATCATACATAGTCGAGCA

6392 CAGCCTCTCTGAGTAGCTGGGATTACAGGCGCCTGCCACCACACTCGGCTAACTTTTTGTA  
TTTCTAGTAGAGACGAAGTTTCAACATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCA  
AGTTATCCACCCACCTCAGCCTCCCAAAGTGGTGGGATTACAGGCGTGAGCCACTGTGCC  
TGGCCTCTAGGATTATATTAATAGAACCAATCTTCAATTATTTTATCTTTCTTTATCTTT  
TTTTCTATGAGAAATGTCTTAAATTTTCAAACCTCAATTTGAAAGCACTTTTAAAT  
[T,C]  
ATACATAGTGAACATTTTATATAAAACAACCTAAAAAGTCTGTGACATTTTGCAGTATA  
AAAATGCAATGSCAGCAGCAGGCTTATTAATTGAGCCTCTTGGAAATGTGGCTGGTCTCT  
AGGTCTGTAGCTCAAAGGCCCTGGCTTGTAACTGCAGGAGCTGACCAGCACAGCTCTAT  
AACCAGSTGTACATCTTCTAGCCTGTGTCCAAGAAAACAGAAATCACAACGCTCTGTGG  
ATAGTGACATCTTAAAGTTTTCTTTCCCTCCCAACTCTTTTGCCAGTTCATTGAATTGCT

9484 GCAACATTTATATCACAATATGTGCTGTTTATGTTCTGAATATCACATATGATTAGTAA  
TCACACAGCTATTTGAGGGCTAAGCATCAGGACTATAAATATTTGATTGTGTTAGTGTCT  
TTGATTGAACCTCTTTTATGTATAATATCTTCAGCTGAATGGGTTTTTATATCAACTTTA  
CTTTTATATAAGCCATGTTTGAATAAACTAGGATTTTAAATCTGAATTTTAAATAGC  
TATGTATGTATCATATATTGTATGCTTTTGAATGTGCTTACCTCTAAGACAAAAAA  
[C,G]  
CTGCTTTCTCTTATTAATTATACATACCATTAATGAATTAGGAAGTTACAGATCACTG  
ATGAATGAGAAATAGGAAAACTTCCCCAATCCACAGTCATAGATCATCTTCTATGAGAG  
AAGAAATGTTCACTTTTAAATGAGGGCTCTATTTTAGGCTTATAAACACTTAGCAGAT  
GAATTTGGTCAGAACCAATTAATCACTAAACATCATGGGGTGTGTTTTGTGTCTAAAT  
AGCCAGACTGGATTAAGCTTTCTCTCTTAAATTTATAGCAAGTGACACAGTATTTTAAAG

FIGURE 3, page 13 of 23

10280 ATAAGAGTGCACATAGCTACAGGCGTTATATAAATTTATAATTCATGGTCCAAATGTACA  
TTTGTATATATGATTTCATTGGGAATTACCAAGGATTAGATCAATTGTGGGSAATATGT  
ATTTTAAAAATAAAACAAAGATAAAGATTCTTTTCTTAATTCAGGTAAAAAGGCAGCA  
TTGCTGCTTCAATTTATAGGTAGATGCTTCTATCAACATTCTTATTTTGTGCTCCAAAT  
CTTGGATTTCGAAAAAATACCAATCCGATATAAATATAAGAAACATACATGCATGTGGG  
[A,G]  
TCTTAACATCAGAAATGACTCTGAATGCACAAAAAAGGGAATTTTC  
GTGCTCCATCTTAGCTTTCTGTGCTTTCTTATTATATATGCAACTGCTGCTGCTGCTA  
TCTTACAAAGTACTTCTGAATCTAATGCACAGGATCAGCAGTAATGCAGCTCAGACTCA  
TCTTTTGGCTTTTGGATTCTTAGATTTCAGATTAAAGTTTAGTCAGGCTATTGAATAGCT  
CTTCAATTTAAAGTCTGATGTGAATATCATGCAAAATATGATGTACATATTCCCATGTG

10297 CTACAGGGGTTATAAATTTATAATTCATGCTCCAAATGTACATTTGTAGTATTGATTTC  
ATTGGGAAATACCAAGGGATTAGATCAATTGTGGGAAAGTGTATTTTAAAAATAAAAC  
AAAGATAAAGATTTTCTTCTGAATTCAGGTAAAAAGGCAGCATTGCTCCTCCATTTAT  
AGGTAGATGCTTCTATCAACATTCTTATTTTGTGCTCCAAATCTTGGATTTCGAAAAAT  
ACCAATCCGATATAAACATAAAGAAACCATACATGCATGTGGGGATCCTAACACCAGAAAT  
[G,A]  
ACTTGTGAATGCACAAAAAAGGGAATTTTCGTGCCCCATCCTTAGCT  
TCTCTGCTTTCTCTATTATATATGCAACTGCTGCTGCTCTATCTTACAAAGTACTTGG  
TAATCTAATGCACAGGATCAGCAGTAATGCAGCTCAGACTGCATGCTTTCGCTTTGGAT  
TCTTAGATTTCAGATTAAAGTTTAGTCAGGCTATTGAATAGCCCTTCAATTTCAAGTCT  
GATGTGAATATCATGCAAAATATGATGTACATATTCCCATGTGCTGAGTAAGTAGATGTAG

10331 AAATGTACATTTGTAGTATTGATTTTCTTGGGAATTACCAAGGGATTAGATCAATTGTGG  
GGAAAGTGTATTTTAAAAATAAACAAGATAAAGATTTTCTGAATTCAGGTAA  
AAGGCAGCATGTGCTCCTCCATTTATAGGTAGATGCTTCTATCAACATTCTTATTTTGT  
GCTCCAAATCTTGGATTTCGAAAAATACCAATCCGATATAAAGAAACATACATG  
CATGTGGGGATCCTAACACCAGAAATGACTCTGAATGCACAAAAAAGGGAATTTTC  
[G,A]  
GGAATTTTCGTGCCCCATCCTTAGCTTTCTGTGCTTTCTCTATTATATATGCAACTGCT  
GCTCCTCTATCTTACAAAGTACTTCTGAATCTAATGCACAGGATCAGCAGTAATGCAGCT  
CAGACTGCATGCTTTTCGCTTTTGGATTCTTAGATTTCAGATTAAAGTTTAGTCAGGCTAT  
TGAATAGCCCTTCAATTTCAAGTCTGATGTGAATATCATGCAAAATATGATGTACATATT  
CCCATGTGCTGAGTAAGTAGATGTAGCATTTGCTAATGTGCTATACATTTAGCATCTAA

10536 TACCAATCCGTATAAACATAAAGAAACCATACATGCATGTGGGGATCCTAACACCAGAAA  
TGACTCTGAATGCACAAAAAAGGGAATTTTCGTGCCCCATCCTTAG  
CTTTCTCTGCTTTCTCTATTATATATGCAACTGCTGCTGCTCTATCTTACAAAGTACTT  
CGTAATCTAATGCACAGGATCAGCAGTAATGCAGCTCAGACTGCATGCTTTTCGCTTTGG  
ATTCTAGATTTCAGATTAAAGTTTAGTCAGGCTATTGAATAGCCCTTCAATTTCAAGTG  
[T,C]  
TGATGTGAATATCATGCAAAATATGATGTACATATTCCCATGTGCTGAGTAAGTAGATGTA  
GCATTTGCTAATGTGCTATACATTTAGCATCTAAGTTATGAACCAGATTCTACCACTGG  
GTAAACATTAAGGGTATGTGTAGAGTTCTGAAAGAAATTTCTCAGCTCCCCAAATCCAC  
ATACTTTTGAAAGCTGATGATTGAAAGATTAATGTGATCCTTTATTGTAACATCTAAC

11548 ACCATTGATTCTTGTTTGGAGAACATTTTGATATATTGCTTATTGGTTTTTGAGGTTGCA  
TCTTTTGGGCTTATAATTTCTATATGATGTTTATTTACATGTTTGAGACTCCAGCATGGA  
ATTATATGACAAAAATATTTTAGTCATTAAACAATCTCTTAACAAGGCTATTTTATCT  
TTGATTGTAGGGTCTTTGATTATGAAAAATAGGAAGAAAGGCATTTCGATGGCCGGGA  
AAAAATTGGAGCTTTTGTTCATTACAATGTGAACATTGGAGGTAAGGGGATATACTTT  
[T,C]  
CAATGGATCCCATAACTTTCTATAGCGTCTTCAATAAATAAGAAACTTATGGCAATAA  
ACAGGCACCTTTAGATACAGAAAAATGCTAATTATAGTTCTTAAATTTTAAATGATAGT  
TTCTTAAATAGGTTTGTGCTGCTTTAATTAAGAAAGCAATATCTAAGAATGAAATAA  
CATATAAAACCTGCCAATTGAATTTAGAAATTAAGATATAAATAAAGCTTTCTTGAT  
TTTAAATGTTATTATAGCATGAATTATTAATTTTAAATTTGAAGAATTTGTGCTTATAT

11917 TTTAGATACAGAAAAATGCTACTTATAGTTCTTAAATTTTAAATGATAGTTTCTTAA  
TAGGTTTGTGCTGCTTTAATTAAGAAAGCAATATCTAAGAATGAAATAACATATAAA

FIGURE 3, page 14 of 23

12840 12844 12847 13019 13022

ACCCTGCCAATTGAATTCTAGAATTAAATATAAAATAAAAGCTTTCTTGATTTTAAATG  
TTATTATAGCATGAATTATTACTCTTAAATTAAGAAATTTGTCTTATATCTGTCTATT  
GA AAAACAGTTGACGTTTTCTATGTGTACTACTTCGATTTAAGTAAAGTAAAAAGTGG  
[G, T]  
TGCTTG333GACATAGCCAAATGCTGTGCTTTAAAGGCGAGGCTGCACTGAGCGAGC  
CCACTAGACAGTGTCTCTGGAAGTTTACAAAGTAAAGGCTGAGGCTAGGCTCAAAATGCA  
CTATAAAGCCCGGTTTTGTTGATTCTATGATTTTAAATTTCCCACTGAATTATGATTTG  
CAGTGTAGGAGCTAGAAATATATATATATATTTAAACAAATGTTCTCTGTTGGTGTGTT  
TGCCCAACAGGCTTCATACGTTTTCTGTGTGTGCTTTG333CTCAGAAGGCATCCAAAAGC  
12840 GACTATGCGAGTAGTCTTCTAACTGGTCTTCTG333TTGAGTTTCCCTGCTCTCAGATA  
AACTCTAAATTTGTTCTCCAGATAAACTTTCTGAAATTTGAGTCTGTTTCTACTTTTGTG3  
TGCATAAAATTTCTTCAGCATGCTTTATTATTCTAAAGGAAAAACTTAAACTCATTGGAG  
TGACACAAGATCTTCTGTCTAGTTCTTCTGCTCAATTTTCTAAACTTTCTAGCAATGCG  
CATATCTATCTATCTTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT  
[T, -]  
ATGATCTATCAATTTATCCATCATCTATACCTAGTGTCTGTTCAAAACATAACAAA  
TTATATTTATTTCCCTAACAGTACTATTTTAAATTTTAAAAATCATCCATGCTCTTT  
TTACAGGCTACTTTCTCCCTTGACTGTCTCTCAAGTCTCTCAACCTAACACACAAGG  
CACACACACACACACACACACACACACACACACACATTTTCTCTCACTCTGCTCACT  
CTGGTCTATTGCTCTCTAGACTGTAATACTAGTTCTCTGGGCTCTCATGGTCTGTTG  
12844 ATTGCGAGTAGTCTTCTAACTGGTCTTCTG333TTGAGTTTCCCTGCTCTCAGATAAACT  
CTAATTTGTTCTCCAGATAAACTTTCTGAAATTTGAGTCTGTTTCTACTTTTGTGCTGCA  
TAAATTTCTTCAGCATGCTTTATTATTCTTCAAGGAAAAACTTAAACTCATTGGAGTAC  
ACAAGATCTTCTGTCTAGTTCTTCTGCTCAATTTTCTAAACTTTCTAGCAATGCGCAT  
TCTATCTATCTTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT  
[A, -]  
TCTATCAATTTATCCATCATCTATACCTAGTGTCTGTTCAAAACATAACAAAATAT  
ATTATTTCCCTAACAGTACTATTTTAAATTTTAAAAATCATCCATGCTCTTTTCA  
CAGGCTACTTTCTCCCTTGACTGTCTCTCAAGTCTCTCAACCTAACACACACGACACA  
CACACACACACACACACACACACACACACACACATTTTCTCTCACTCTGCTCACTGCT  
TCTATTGCTCTCTAGACTGGTAAATACTAGTTCTCTGGGCTCTCATGGTCTGTTGATC  
12847 GCAGTAGTCTTCTAACTGGTCTTCTG333TTGAGTTTCCCTGCTCTCAGATAAACTCTA  
ATTGTTCTCCAGATAAACTTTCTGAAATTTGAGTCTGTTTCTACTTTTGTGCTGCA  
AATTTCTTCAGCATGCTTTATTATTCTTCAAGGAAAAACTTAAACTCATTGGAGTACACA  
AGATCTTCTGTCTAGTTCTTCTGCTCAATTTTCTAAACTTTCTAGCAATGCGCATATCT  
ATCTATCTTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT  
[T, -]  
ATCAATTTATCCATCATCTATACCTAGTGTCTGTTCAAAACATAACAAAATATATT  
TATTTCCCTAACAGTACTATTTTAAATTTTAAAAATCATCCATGCTCTTTTCA  
GCTACTTTCTCCCTTGACTGTCTCTCAAGTCTCTCAACCTAACACACACGACACAC  
ACACACACACACACACACACACACACACACACATTTTCTCTCACTCTGCTCACTGCT  
ATTGCTCTCTAGACTGGTAAATACTAGTTCTCTGGGCTCTCATGGTCTGTTGATC  
13019 CTGACACAAGATCTTCTGTCTAGTTCTTCTGCTCAATTTTCTAAACTTTCTAGCAATGCG  
CATATCTATCTATCTTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT  
CTATCTATCTATCAATTTATCCATCATCTATACCTAGTGTCTGTTCAAAACATAACA  
AATTATATTTATTTCCCTAACAGTACTATTTTAAATTTTAAAAATCATCCATGCTCTT  
TTTTACAGGCTACTTTCTCCCTTGACTGTCTCTCAAGTCTCTCAACCTAACACACA  
[C, -]  
GCACACACACACACACACACACACACACACACACACATTTTCTCTCACTCTGCTCA  
CTGGTCTATTGCTCTCTAGACTGTAATACTAGTTCTCTGGGCTCTCATGGTCTG  
TTTGTATCTAGTATGTTACTGTTTCTTAAAGGATTTTAAAAACCTTGAGTAGAGAATA  
AGCTTTTGGAGTCTGATGGACCTGAATTTGAGTGTGTTTCTGTCACTATCTGTGAACCTG  
GGAAGATCACTGTACTCTTTGCTGATTTTCTATGATATAAAATACCTTACAAAAGC  
13022 ACACAAGATCTTCTGTCTAGTTCTTCTGCTCAATTTTCTAAACTTTCTAGCAATGCGCA  
TATCTATCTATCTTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA  
TCTATCTATCAATTTATCCATCATCTATACCTAGTGTCTGTTCAAAACATAACA  
TATATTTATTTCCCTAACAGTACTATTTTAAATTTTAAAAATCATCCATGCTCTTT  
TTACAGGCTACTTTCTCCCTTGACTGTCTCTCAAGTCTCTCAACCTAACACACAAGC

FIGURE 3, page 15 of 23



[illegible]

18528 ACAAGAGATGTGTCAATATCTGTCTTTGTTGTCTTATTGTACAATAGAGTCACTGGGT  
AGAAAATCTTGTCTTCCAGCTGATGGTCTATGGTTCATTTGTATTCTTTCTCTTGA  
AGTTGTTGATATTGCTTGGGAACAAAGGATATGAACCTATTATAGCTGTTTCTCTTT  
CTTTAAGGGAGGATATTATATAATAATTCTCAACTCTTTAATCTAGACATCAGTAAAC  
TCAGTCTTCAATCTCACTAAATAGCAAAACTTTCCCCATAAATTCTGATTTAGCTCATAA  
[S, A]  
AAATTTCAGAACACTTTCAAGTATTTTGATGTCTTTGATTTACTTTGAAAATTACATGTA  
GAGTTACTTCAGAAGCCTGACAATTGATCTTTGGCAGCCAGGTTCCCTCTTAAGTGGTT  
TTCAGAAGCTTTTCAGGTAGTCTGGACCTCTGGCAGTAGTACTTTGCTGACTCTATAGG  
TTCTTTTCTCATTTAAAGTCATCTCATTATGAAATGCAAAAGCTTTCTATGTTAAGAG  
CTGTTTCATCTTTATGTTAATTATATTCTTATTCAGTGGGCAAGCTTACTGACCTACGTG

18722 TATTATATATAATAATTTTCAACTCTCTTTAACTGTAGACATCAGTAACTTCTAGCTCTCATTCT  
CACTAAATAGCAAAACTTTCCCATATAAATTTCTGATTTACCTCATAAAAAATTTGACAACA  
CTTTCAAGTAATTTTGATGTCTTTGATTTACTTTGAAAAATTACATGTAGAGTTAGCTCTAG  
AAGCCTGACAATTTGATCTTTGGCAGCCAGGTTCTCTCTAGAATGGTTTTCAGAGGTCTTT  
CAGGTAGCTCTGAGCTCTCTGGCAGTATGACTTTTGTGACTCTACTAGGTTCTTTCTCTCAT  
[T, C]  
TAAAGTCATCTCATTATGAAATGCAAAAGCTTTCTATGTTATGGAGGCTTTTCTATTTT  
TCTTAATTTATATTCTTATTCAGTGGGCAAGCTTACTGACCTACGTGAAATGAGATCTCTCT  
TCTCTAGAGGAAATGATTGTTTTTAAAGACTGAAGGACTAGTCTTTTAAAGAAATGCAAAAT  
GAATCTCTCATTAGCTCTCTAAGACAAATTTAAATCAGCTATAAGTTTATGTAATAAATAT  
GTCTCTATGATTAGCAATATAGATATACCTTTTTTATTATTATTTTCATTCTTGAAGAGTCT

18775 TCAATTCCTCACTAAATAGCAAAACTTTCCCATATAAATTTCTGATTTACCTCATAAAAAATTT  
CAGAACACTTTCAAGTATTTTGATGTCTTTGATTTACTTTGAAATATCATGTAGAGGCT  
ATCTCAGAGGCTGACAATTTGATCTTTGGCAGCCAGGTTCTCTAGAATGGTTCTCA  
AGCTTTTCAGGTAGTCTGGAGCTCTGGCAGTAGTACTTTGCTGACTCTACTAATTTCTCT  
TCTCATTTAAAGTCATCTCATTATGAAATGCAAAAGCTTTCTATGTTATGAGGCTTTCT  
[C, G]  
ATCTTTATGTTAATTATATTCTTTATTCAGTGGGCAAGCTTACTGACCTATCTGAAATGAG  
CTGTTCTCTCTTAGGGAATGATTGTTTTTAAAGACTGAAGGACTAGTCTTTTAAAGAAAA  
TGAATGAATCTCATTAGCTCTCTAAGACAAATTTAAATCAGCTATAAGTTTATGTA  
TAAATATGTCTCTCATGATTAGCAATATAGATATACCTTTTTTATTATTATTTTCATTCTGA  
AAAGTGATTTTTTTTTTGAAGTTTAAAAACAAAGCTTGGTGTTCTTTCTCTCTGAGTC

18951 CAGAAGCTTTTTCAGGTAGTCTGGAGCTCTGGCAGTAGTACTTTGCTGACTCTACTAGGTT  
CTTTTCTCATTTAAAGTCATCTCATTATGAAATGCAAAAGCTTTCTATGTTAGGAGGCT  
GTTTCATCTTTATGTTAATTATATTCTTATTCAGTGGGCAAGCTTACTGAGCTAGCTGAA  
ATGAGCTCTCTCTCTTAGGGAATGATTGTTTTTAAAGACTGAGGCTATGTTTAAAG  
AAAAATGGAATGAATCTCATTAGCTCTCTAAGACAAATTTAAATCAGCTATAAGTTTA  
[T, C]  
GTACTAAATATGTCTCTCATGATTAGCAATATAGATATACCTTTTTTATTATTATTTTCATT  
TTGAAAGTGATTTTTTTTTTGAAGTTTAAAAACAAAGCTTGGTGTTCTCTCTTTTCC  
AGTCGGTCCCGGAGAAAAATGCAACGGTGTCAAATATTTCCATCACGGGGATGCTTTGTC  
ATGTACCTGCTTGGCGCCCTCTTGGTTACCTAACCTTCTATGTTAGGTGACTCTGAAAG  
TCATTCTCTATATGCAAACTCTTGTAGGCTGGTCTTGACCTGGGTAGGTATGATTTTT

18974 ACTCCTGGCAGTAGTACTTTGCTGACTCTACTAGGTTCTTTTCTCATTTAAAGTCATCT  
CATTATGAAATGCAAAAGCTTTCTATGTTAGGAGGCTGTTTCATCTTTATGTTAATTATA  
TTCTTATTCAGTGGGCAAGCTTACTGACCTACGTGAAATAGACTGTTCTCTCTAGGGA  
AATGATTGTTTTTAAAGACTGAAGGACTAGTGTTTAAGAAAAATGGAATGAATCTCATT  
AGCTCTCTAAGACAAATTTAAATCAGCTATAAGTTTATGTACTAAATATGCTCTCATGAT  
[T, G]  
AGCAATATAGATATACCTTTTTTATTATTATTTTCATTTTGAAGTGATTTTCTTTTGA  
AGTTTAAAAACAAAGCTTGGTGTTCTTTCTTTTCCAGTCTGGTCCCGGAGAAAAATGCA  
AACGGTGTCAAATATTTCCATCACGGGGATGCTTGTCTATGACTGCTTGGGCGCTCTT  
TGGTTACCTAACCTTCTATGTTAGGTGACTCTGAAAGTCATTCTCTATATGAAATCTCT  
GTTAGGCTGGTCTTGACCTGGGTAGGTATGATTTTTTAAAAATTGCTTCTATAGGATG

19540 GGTATGATTTTTTAAAAATTGCTCTCTATAAGCATGCTCTATAGATGACACATATTTCAAT  
AATATACTATTTTAGTTTTGTCACTTGACCTGAGGAAATGGGGCTGATTCAGCTGGCT  
AACAAGTTACAGAATTTTGAAATTAACACCTATTTTAAAAAAATATCTCTCAAAACAAA  
ATTATTTTCTCTAGGGATAGATGATTTTCTCTGGCTAGACTCCATAGTCTCAACTCAGG  
CTACAGTGATGAGAATGAATCCACTTGCATGTGATAAGGCTCCTTTGATGGAATATTA  
[A, C]  
CTGCCACACAAATAGCAGGGAAGCTGCCAGGTCTCAAGTTGAATTTGCTCTCTCTTTA  
CCAGTCAAGTCAAATCTGGGAGCTTGGGACTTTAGGTAAAAATTTCTGACATATCTCATTC  
TATTTTGTATTACTAAATGATTTCTAAGAAAGAGGACATGACAGAATTTCTCTCAATCT  
AAGAAATGCACCACCAAAAAAAGTGAATGAGGACATTAGATTATGCTCTGCAACATTT  
CTCTCTGGCATCTTAACAGTTCAAAAGGGAGTAGGATTGTACTCTCTCTATGAAAGTGTG

19841 CTGCCACACAAATAGCAGGGAAGCTGCCAGGTCTCAAGTTTGAATTTGCTCTCTCTTTA  
CCAGTCAAGTCAAATCTGGGAGCTTGGGACTTTAGGTAAAAATTTCTGACATATCTCATTC

FIGURE 3, page 18 of 23





[G, T]

GCATATGAAGACTATTTCTATAATGACACTGTGGTCACTTTAAATGCAGCTTGTGTGCTG  
AAATATATTTTGGCACAATTCCTTTTCATGASTGATGAAATCAGATCTGTACTACATG  
GTGGCTAATATTTTACTCTTAAATCATGTCTTGGCTCTAATAATCTGAAAGTATTTAG  
ATGACATACACATAGTTTAAAGCTAAATGAGTCCCTCTTGGGTACAAAGACAGAAACACA  
ACTATAAACAGAGGTATACATAGAGGGTAAATTTGCCAGCAACAACTTCACTGAGAAA

22783

TGAGAAATAAAGCACTGATATAAATCTGACCAAGCAACAGCAATACTGTGTAAACATT  
AGATGCCATTAGAACTAAATTTGACCAATAAGCAAGAGTTCAGAAAAATGACTAACTGC  
TGTCCTTCATTATGTATTTCACTCAACATTAGCAATTTATGAAACATTTTGCACATATC  
CTGTCCCTCAGCTTGGCAATTTACATTTATATAATCTGTGTAAGTGTCTCACTGCCCCAC  
AGAGTCTATAAGTCCCTGGGACTTGGTATGTGTACAGTGTACTGGCACAGAGGTGTAGCTC  
[C, T]  
GTGGTCTGTGGGAGAAAAATGCTCTTCAAAATGAATCTTGCCTTGTCTTGAATGTATAA  
ACTGCTTTTCTAGCAAAAGCATAGACTACTCTTCCCTTGGTGACATGTCTACGAATTC  
AGCTGGGTGAGGATCTGGCTTAAATGAACCAAACTCCCTATACATGAAGCATACACAG  
AGATGTGACAGAGAGTGGTCACTTCCCTGAGTGGATCTCAATCAAGTCTCTGAAGCTA  
AATTCATTTTCTTCTTAAATAAATAAAAGTTGTTATTGGCGCTTTTGTGTGTTT

22787

AAATAAAGCACTGATATAAATCTGACCAAGCAACAGCAATAAGTGTGTAAACATTAGAT  
GCCATTAGAACTAAATTTGACCAATAAGCAAGAGTTCAGAAAAATGACTAACTGTGTCT  
CTTCATTATGTATTTCACTCAACATTAGCAATTTATGAAACATTTTGCACATATCTCTGT  
CCTCACCTTGCATGTTTACATTTATATAATCTGTGTAAGTGTCTCACTGCCCCACAGAG  
TCATAAGTCCCTGGGACTTGGTATGTGTACAGTGTACTGGCACAGAGGTGTAGCTCTGT  
[G, A]  
TGCTTGGGAGAAAAATGGTCTTCAAAATGAATCTTGCCTTGTCTTGAATGTATAAACTG  
CCTTTTCTAGCAAAAGCATAGACTACTCTTCCCTTGGTGACATGTGTCTACGAATTCAGCT  
GGGTGAGGATCTGGGCTAAATGAACCAAACTCCCTATACATGAAGCATACACAGAGAT  
GGTACAGAGAGTGTCTACTTCCCTGAGTGGATCTCAATCAAGTCTCTTGAAGCTAAATTT  
CAATTTTCTTCTTAAATGATAAAAGTTGTTATTGGCGCTTTTGTGTGTTTATT

22825

CAATAGTGTGTAAACATTAGATGCACTTAGAACCAGCAATAGTGTGTAAACATTAGAT  
AGAAAAATGACTAACTGTGTCTCTCATTTATGTATTTCACTCAACATTAGCATTATGA  
AACATTTTGCACATTATCTGTCTCTCACCTTGCATGTTACATTTATATAATCTGTGTA  
AGTGTCTCACTGCCCCACAGAGTGTAAAGTCCCTGGGACTTGGTGTGTGTACAGTGTACT  
GGCACAGAGGTGAGCTCTGTGTCTTGGGAGAAAAATGGTCTTCAATGAATCTTGC  
[T, C]  
TTGTCTTGAATGTATAAACTGCTTTTCTAGCAAAAGCATAGACTCTTTCCCTTGGT  
GACATGTGTCTACGAATTCAGCTGGGTGAGGATCTGGGCTAAATGAACCAAACTCCCTA  
TACATGAAGGATACACAGAGATGGTGACAGAGAGTGGTCACTTCCGTGAGTGGATCTCAA  
TCAAGTCTCTGAAGCTAAATTTCAATTTTTTCTTACTAAAAATGATAAAAGTTGTTAT  
TGGCGCTTTTGTCTTATTTCTGTATAACTTAGGGCTCAGATTTTCAATGTGTCAAATG

22967

CCTCACCTTGCATGTTTACATTTATATAATCTGTGTAAGTGTCTCACTGCCCCACAGAG  
TCATAAGTCCCTGGGACTTGGTGTGTGTACAGTGTACTGGCACAGAGGTGTAGCTCTGT  
GTGCTTGGGAGAAAAATGGTCTTCAAAATGAATCTTGCCTTGTCTTGAATGTATAAACT  
GCCTTTTCTAGCAAAAGCATAGACTCTTTCCCTTGGTGACATGTGTCTACGAATTCAGC  
TGGGTTGAGGATCTGGGCTAAATGAACCAAACTCCCTATACATGAAGGATACACAGAGA  
[A, T]  
GGTGACAGAGAGTGGTCACTTCCCTGAGTGGATCTCAATCAAGTCTCTGAAGCTAAATTT  
CAATTTTTTCTTCTTAAATGATAAAAGTTGTTATTGGCGCTTTTGTGTGTTTATT  
CGTATAACTTAGGGCTCAGATTTTCAATGTGTCAAATGCTGACTCACAGCATGGTCTCTC  
TGACAGTTTATTTCAATTTAAGCAACTCTTACCAGTAAGTTATTTACTTGCCTTGATAT  
CTCCACACATTAATAATAAACTAACAAAACCTAATCTGAATTAATCTATCAGCTTTA

23248

CATGAAGGATACACAGAGATGGTGACAGAGAGTGGTCACTTCCGTGAGTGGATCTCAATC  
AAGTCTCTGAAGCTAAATTTCAATTTTTTCTTCTTAAATGATAAAAGTTGTTATTG  
GCGCTTTGCTTGTATTCTGTATAACTTAGGGCTCAGATTTTCAATGTGTCAAATGCT  
GACTCACAGCATGGTCTCTCTGACAGTCTTATTTCAATTTAAGCAACTCTTCAACAGTAAGT  
TTATTTACTTGCCTTGATATCTCCACACATTAATAATAAACTAACAAAACCTAATCTGA  
[A, G]  
TTAAATCTATCAGCTTTAGGCATTATTTGTGTCTCTCTTCTTCAACATGCTAACTGG  
GCTCTCTTCTTAGGAGCTTGAGAAGATATGACTGGGGTTGTTTTCTCTACTTCATT

FIGURE 3, page 20 of 23

23764 GAAATGAATTAAGGTAGTCTTTTCACCTAGCAGCCCAACATTTCCTTCTCTCTTTAGTAGTGTG  
GGTGGGATATCAGTTATGCAATGGACCTCTCTCCACAGGAATGACCATGTTCATTTTCAG  
CTTATGCTTCCCTTTATGCAGTAAAGTTCCATATTTCCATAAAGAACAAAGAAACCAAAAT  
AATCTTAATGGATATATAAGCAACACACAGATGAAAAATTCACCTGCCACGCTTTGAAA  
AAAGATCCCTAGCTACTGTAATTCATCTATAAATAAAATCACTCTTTTCACTTATGTT  
[G,T]  
TCTTCAGATCTCCTGTTTGAAGTTATATAGATATCAACATAGAAATGACGCGTATATT  
GCTATCACTGCAGTGGAGCAGTGAATCTGTAGTCTTTCCAAATCTCTTGCTTAAGCAAA  
CCTGCAAAAATCAAAGTGAGAGTAAAGTTAAACAATGGGAGAGCTTTTCTTTTCTTTT  
AAGAGTTAGAAGTAAGATCTCTACTTCTCTGTGCTCCACATTTTGAAGCTTCACAT  
GGGCCCCGTCATCAGAATACAGCAACCCCTAACAGGCTCTCTGTTGAGGAGCTTTCTCTG

23765 AAATGAATTAAGGTAGTGTCTTTCACATAGACGCCCAACATTTCTTTCTCTCTTAGTAGTAGGG  
GTGGGGTATCAGTTATGGAATGGACCTCTCTCCAGAGGACTGATCATGTGATTTTTCAGC  
TTATGCTTCCCTTTATGACATAAAGTTTCATATTTTCATATAAGAACAAGAACCATAA  
ATCCTAATGGATATATAATGAACATACACATGAAAATTTACCTGACATGCTTTTGAAAA  
AAGATCCCTAGCTACTTGTATTTCATCTTATAATTAATAACAGTCTTTTCACTTATGTT  
{C,T}  
CTTCAGATCTCCTGTTTTGAAAGTATATAGATATCAACATAGAAATGCAGCGTATATTG  
CTATCAACTGCAGTGGAGCAGTGAATCTGTAGGTTTTCCAACATCTTTCTTTAAGCAAAC  
CTGCAAAATCAAAAGTGTGACCTAUGTCTAAACAATGGGAGAGCTTTTTTTTTTTTTTA  
AGAGTTAGAACTAAGACTCTCACTCTCTCTGTGGCTCCACATTTTTGACCTTCACATTG  
GGCCCTGCATCAGAATACAGCACCCCTTAACAGGCTCTGTTCAGAGCTCTTCTCTGG

24422 GGAATGCACTGGAACCACCTTCAGTTCTGTTTGAATTTTCTATTCCCTATTTCAGAAA  
GAGGAAGAAGCTTTTGCAATTTACTCCAACCGTTCTACCTATTATTCCCATAAACTTTCTG  
TGATCTCATATCATTTAGGCCAAATGTTAATCTTTCTGGGAGCCAGGAGACTGCTTTCCAA  
TTCAGAGGCCCTGGACATATAAGCTGCCTCTAACTACTCTAACTGACCTATTGACTT  
GAATGCACCTTTTAAACAAGTGACTAAAAACAACACTGTGACTATTCTCTGAAAATGAGC

24538 GATGAGGCTTTGGTGGAATCAGCCAGTGGTCTGCCCTAGCAAAGGTAACAGAACTGCTG  
GGGGCTTTTGGTCTAGGCTCACTACTCAGGGAGGCACCTTAACTAGGAATGACCAGCAA  
GTTTCCTTCTGATCTTTTCCACCACCACCACAAGCCTAGTACCTCCCTCCCTCTTTGCT  
CTGTTGCTCTCTTCGGGAATGCACTGGAACCACCTTCAGTTCTCTTTGGAATTTTCTTA  
TTCCTATTTCAGAAAGAGGAAGAAGCTTTTGCATTACTCCAACCTTCTACCTATTATT  
[C,G]  
CCATAAACTTTCTGTGATCTCATATCATTAGGCCAAATGTTAATCTTTCTGGGAGCCAGG  
AGACTGCTTTACACTCAGAGGCCCTGGACATATAGGACTGCCTCTAACTCACTCTAACT  
CAGCTTATTGACTTGAATGCACCTTTTTAAAGTGAATAAAAACTGTGACTATT  
CTCTGAAATGAGCCTATATCTCATACTATTTATTCTGTTTAACTGTGAACAACT  
AAGTCTCTGGCACTATGTATATACCATAAAAAGCTTATTGTAAGCCTACTAATTGGAC

FIGURE 3, page 21 of 23

24819 AACCGTTCTACCTATTATTCTTCCATAAACTTTCTGTGATCTCATATCATTTAGTSCCAAATGT  
TAATCTTTTCTGGGAGCCAGGAGACTGCTTTTACATTTCAGAGGCGCTTGACATATAGGACT  
GGCTCTAACTCACTCTAACTCAGCTTATTGACTTGAATGCACTTTTAAACAACTGACTA  
AAAAACAACTGTGACTATTCTCTGAAAATGAGCTATATCTGATAATTATTTTATTCTGT  
TTAACACTGTGAAAACAAATTAACTCTCTGTGCACTATGTATATACCATAAAAAGCTTATT  
[C, T]  
GTAAGCCTACTAATTGGACCGAGTTTGTGACAATATGSAATAAGCACTAATTGCAGATCATA  
ATGTAGAATTATAGGCTGCTGAGGAAAAACAATATCACACCATTTCTCTCTGAGTTTCC  
TTTTGAGAATGAGTTTCATAATGTTCACCTAATCCAATTTTTAAATTCCTTACAAAGTTA  
TTCTTAACTATTTTCCAGAGACTATCTGCTTTGTGATTCTAGAAATGAAATGGCTTTTC  
AGCTAAACAGATGGCTTAATTTTGTGTGAGTGTGTATGAAAGCAATGTACATGAGA

25743 TATCCAGTTACAGCAGCTAACTTGAGCAGCTGCTGCAAACTGAGCTTCTTTGACCTT  
CGCTACTTATTTGAGCTGCTAAAAATAGGGCTGAAATCTGTCAAGGACCTCTGAAGGGAG  
GATAAGATTCTACTATTCAATTTAATTTAAGCTTTTATTGAGTGCTGCTGTGTGCACA  
ACACTAAGCTAGAAAGTCTGAGGAATGTTTAGATTATTAGGCTCTCTCTGCTGCTTTCA  
TAGATTTACAATCTATTGATAGGGAGAGCTAAAAAGGAGCAAAAGGGAAGGAGCAAAACA  
[C, T]  
AAAAACGTCAAAAATTTTAAAAATACCATTTTAAAAATTTTATTTAAAAATGTTAAATACCAT  
GCAAAATTAAGGAAAACTTAGATTTCATAAAAAATTCCTTTCACTAATCTGTGTAAATCAAT  
TCAGTGCTTGCCTTAATGTCTCATCCAGTCTGATGAGACATCTTTCTGTATCAACAAGG  
GTTTTACTATGTTTCTTAATATGTGTCTTGCCTGTTATCTCTCTCTGAGGAGATTATT  
TTTAAACAATAAATCTGAAAACCTAAGAAAGTGAAGCATAAAAATCTCTCTTATAAAAAA

26044 AAAACGTCAAAAATTTTAAAAATACCATTTTAAAAATTTTATTTTAAAAATGTTAAATACCAT  
GCAAAATTAAGGAAAACTTAGATTTCATAAAAAATTCCTTTCACTAATCTGTGTAAATCAAT  
TCAGTGCTTGCCTTAATGTCTCATCCAGTCTGATGAGACATCTTTCTGTATCAACAAGG  
CTTTACTATGTTTCTTAATATGTGTCTTGCCTGTTATCTCTCTCTGAGGAGATTATT  
TTTAAACAATAAATCTGAAAACCTAAGAAAGTGAAGCATAAAAATCTCTCTTATAAAAAA  
[G, C]  
GCCAAGGAAAAATGACACTCCATTTTCAAATATCAAAGTTAGCATCAAGACTGCACAAG  
ATGAATGTACAGTCATGTGTTGCTTACAAATGTGGACATATCTGTGAGAAATGCATCTTTA  
GGCAATTTTGTCAATGTGCAACACCATAGATTGTACTTGCAGCTAATTTGGTGGAGCCT  
ACTATACACTAAGGCTATATGGCATAGCCTAGTACTCTTAGGCTACAAACCTGTACAGCA  
TGTTACTGTACTGAATAGTGGAGGTACCTGTAAACATAATGGTAAGTATTGTGTCTCCAA

26555 AGTACTCCTAGGCTACAAACCTGTACAGCATGTTACTGTAAGTGTGAGGTACCTG  
TAACATAATGGTAAGTATTTGTGTCTCCAAACGTAGAAAAGCTACTGTAAAAATACAGTA  
TTACAACCTTAGGGTATCACTGTCTTATATGTGGTCTGTTCTTGAAGTGAATGACTATGC  
TTAATACCACTGAAGTGTACACTTAAAAATGGTTAAGATGTAATCTATGTTATGTAT  
GTTTATAATAATAAAAAAATGAAAAAAGCATCAACATCTTTCTGGGAAAAAAGAAAA  
[G, A]  
GAAAGAAAATGCATTAGAGTGATGAGAATATTTGAAGTAATAGATAAAGTCAAAAAACAA  
GAAATGATCTTGCCTTTGAACCTTTCTGTGTTAAGATTTCGTACATCAGTGATCACACTGTT  
ATTTCCCAAACGACCCTTCAGCTGGATACGACATTTCTGATTTCAGCTGTGCTTATTGC  
ACTTAATAATGTTCTGGTCATCCTTGTGCCAACTATAAAATACATCTTCGGATTCAATAGG  
TGAGTTTCAGAAAGGCTTCAATTTGGTCAACCCAACTCAGGCTCATTAAATGATGGAC

27886 GGTTTTATTAAAGTGTGTGCTGGCATCTCCTTTGCTAGGAAGTGTGGTAAGACATTGA  
CCTTGGCCTGTGTTTGTCTTCTCAGGGGCTTCTTCTGCCACTATGCTGATTTTTATTCTT  
CCAGCAGTTTTTTATCTTAACTTGTCAAGAAAGAAACTTTTAGGTCACCCCAAAAGGTC  
GGGGTAAGTAAACCTTGCAATTTCCCCCATATTAGTTGTCTTCCAACTACTTAGAATA  
AACTAGAAAAATACACATAGTTTCAGAAAAATGAATCAATGTACAAAGAAACAAAAATCAAAA  
[A, C]  
TGGGCTAGAACTTTCTGGTAGCAGAGAAAGGGGACATATTTCTGAAACTCAAAATGATTCT  
ACTTCAAATATCAAATATCTGTGTTGAGTCTGTACATATGTCAAATAGTAGTAGGCTT  
TCCACAGACACATATGCTTCAGGCAAAATAGCAGTGTCCAATAACCAAGCTGTGTTGTGC  
TATCCGTGAAAAATCATGCAAGAAGGAATTAGGCTCCCTAGGCTGTATTAGGAATAATTT  
AAATATTTTGGTCATGGTTGTTAGGTTTGCAAGGCCAAAGGAAAGATGTTCCTTTGTGTT

31884 CTTTTATGGTTAGTTTGAAAGAAATCCATTGAAGATAGAAAATGAGAGAATAGAAAGAAAC  
TGAGAATAGTAAAAATAAGAGCAGAGAAAATATGGGGGCAGGCAAAATATGTGAGTGCTA

ATTTTAGCCCTTCTAGTGGGCTCTGAAATGGTAACCTATTGTGATTTCATTTCCTCTTCT  
GTGACCACTAATGTTGAAACCTTTCAAGGTGTTAATGGTCACTCATATATCTCTTTTG  
TGAAGTGTCTATTCAAATCTTTTGCCCATTTTAAATATTTAGTGTATGPGTTTATTTGG  
GTATTTGTAGAAGCTCTTTAAATATGGATCCATGTCAGATGTCAGATATATTTTCCAG  
TCTATGGTATGGTTGCTTATTTTCTAAAGGTGTCTTAATTAATCTTCTGGGGGCGAGG

TTTCATTTCCTGCTCTCTGTGACAACTAATGTTGAAAACCTTTTCAAGTCTTTAATGCTCACT  
CATATATCTTCTTTTGTGAAGTGTGTAATCAAATCTTTTGCCATTTTTAAAAATTTAGGT  
TATGTGTTTTTATTGGGTATTGTAGAAGCTCTTTAAATATGGATCCATGTCCAGATTGC  
CAATATATTTTCCAGTCTATGGTATGGTTGCTTATTTTCTTAAAGGTGTCTTAATTACA  
TCTTTCTGGGGCCAGTCCACCATAGCTCAAAGTTTGAATTTATGTCTTAATGAGATAA  
[T, A]

ATTAAATCAGAGTGGTATAGTCAAATTAATGTTTTGATGTCCTGGGCCATATAGGTAG  
GACTGGATCATCTAACCAAGATGCAAAAAAAAAAAAAACAAAAACAATACTACTTG  
GAAAACTTATTTTAAATTAACA

[illegible]